PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/705, A61K 38/17, 35/00, C07K 16/28, C12N 15/11, A61K 39/70, 39/395, C12Q 1/68, G01N 33/68

A2

(11) International Publication Number: WO 00/03015

(43) International Publication Date: 20 January 2000 (20.01.00)

(21) International Application Number: PCT/US99/15482

(22) International Filing Date: 8 July 1999 (08.07.99)

(30) Priority Data:
09/113,427 10 July 1998 (10.07.98) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/113,427 (CIP)
Filed on 10 July 1998 (10.07.98)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). REDDY, Roopa, M. [IN/US]; 1233 W. McKinley Drive, Sunnyvale, CA 94086 (US). GORGONE, Gina, A. [US/US]; 777 San Antonio Road #6, Palo Alto, CA 94034 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). AZIMZAI, Yalda [US/US]; 213 Famoso Plaza, Union City, CA 94587 (US). PATTERSON, Chandra [US/US]; 2189 Leland Avenue, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: HUMAN TRANSPORT PROTEIN HOMOLOGS

(57) Abstract

The invention provides a human transport protein homologs (HTPH) and polynucleotides which identify and encode HTPH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or preventing disorders associated with expression of HTPH.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# HUMAN TRANSPORT PROTEIN HOMOLOGS

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human transport protein homologs and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, reproductive disorders, and copper metabolism disorders.

## BACKGROUND OF THE INVENTION

Eukaryotic cells are bound by a lipid bilayer membrane and subdivided into functionally distinct, membrane bound compartments. The membranes maintain the essential differences between the cytosol, the extracellular environment, and the contents of each intracellular organelle. As lipid membranes are highly impermeable to most polar molecules, transport of essential nutrients, certain metal ions, metabolic waste products, cell signaling molecules, macromolecules and proteins across lipid membranes and between organelles must be mediated by a variety of transport molecules.

Transport across membranes depends on transporters, or pumps, which are membrane-spanning proteins that bind to specific classes of molecules and undergo a series of conformational changes in order to transfer the bound molecule across a membrane. Transport can occur by a passive, concentration-dependent mechanism or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Examples include facilitative transporters, the secondary active symporters and antiporters driven by ion gradients, and active ATP binding cassette transporters involved in multiple-drug resistance and targeting of antigenic peptides to MHC Class I molecules. Transported substrates range from nutrients and ions to a broad variety of drugs, peptides and proteins.

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport small hydrophilic molecules across biological membranes. ABC transporters are comprised of two homologous monomers, each containing two parts: a transmembrane domain with multiple transmembrane segments and a nucleotide binding domain. Mammalian ABC transporters are found either as dimeric transporters, e.g. the multiple drug resistance transporter and the cystic fibrosis

transmembrane regulator proteins, or as monomeric transporters, e.g., the transporters associated with antigen processing; TAP1 and TAP2 proteins, which dimerize to form the complete active TAP transporter. Two monomeric ABC transporters have been identified in the human peroxisome membrane: the adrenoleukodystrophy protein (ALDP) and the 70-kDa peroxisomal membrane protein (PMP70). Mutations in the adrenoleukodystrophy gene cause X-linked adrenoleukodystrophy, an inborn error of peroxisomal β-oxidation of very long chain fatty acids. Mutations in the PMP70 genes have been found in patients with Zellweger syndrome, an inborn error of peroxisome biogenesis. Multidrug-resistance (MDR) results from overproduction of another member of the ABC transporter family, P-glycoprotein. MDR is primarily caused by increased drug extrusion from the resistant cells by P-glycoprotein. The P-glycoproteins have 2 homologous halves, each with 6 hydrophobic segments adjacent to a consensus sequence for nucleotide binding. The hydrophobic segments may form a membrane channel, whereas the nucleotide binding site may be involved in providing energy for drug transport. (Saurin, W. et al. (1994) Mol. Microbiol. 12:993-1004; Shani, N., et al. (1996) J. Biol. Chem. 271:8725-8730; and Koster, W., and Bohm, B. (1992) Mol. & Gen. Genet. 232:399-407.)

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states. In particular, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. A deficiency in copper may result in the hepatolenticular degeneration of Wilson's disease, as well as anemia, growth retardation, defective keratinization and pigmentation of hair, hypothermia, degenerative changes in aortic elastin, mental deteriorations, and scurvy-like

changes in the skeleton. An excess of copper may produce hepatitis, cirrhosis, tremor, mental deterioration, hemolytic anemia, and renal dysfunction. Menke's disease is an X-linked recessive disorder of copper metabolism, wherein copper uptake from the intestine is normal, but the tissue distribution of copper is perturbed. In particular, the content and activities of copper-containing enzymes including ceruloplasmin and lysyl oxidase are decreased. Symptoms of Menke's disease include decreased amounts of collagen and elastin, which produce aneurysms, sudden cardiac rupture, emphysema, and osteoporosis. Typically, death occurs within five years. (Isselbacher, K.J., et al. (1994) Harrison's Principles of Internal Medicine, 13th edition, McGraw-Hill, Inc. New York, NY., pp. 481-482; Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, 5th edition, W.B. Saunders Co., Philadelphia, PA, pp. 134-135, 863-864.)

The energy requirements of most mammalian cells are met through a continuous supply of glucose which circulates in the blood. Glucose enters cells through specific glucose transporter molecules present in the plasma membrane. The family of glucose transporters includes passive transporters typical of mammalian tissues and active, $H^+$-linked sugar transporters from bacteria. These transporters characteristically contain two groups of six putative, membrane-spanning, alpha-helices separated by large, hydrophilic, cytoplasmic regions. Both the N-terminal and C-terminal regions of the sequence are also predicted to be cytoplasmic. Biophysical studies on the human erythrocyte glucose transporter indicate that the membrane-spanning alpha-helices associate to form a hydrophilic channel or a substrate-binding cleft extending across the membrane. The mechanism of substrate translocation involves alternate exposure of the substrate-binding site to each face of the membrane via a conformational change. (Pessin, J. E. and Bell G. I. (1992) Annu Rev Physiol 54:911-930.)

The discovery of new human transport protein homologs and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cancer, reproductive disorders, and copper metabolism disorders.

## SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human transport protein

homologs, referred to collectively as "HTPH" and individually as "HTPH-1," "HTPH-2," and "HTPH-3." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and fragments thereof, as well as an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, and fragments thereof.

The invention further provides an expression vector containing at least a fragment

of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof, as well as a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

The invention also provides a method for treating or preventing a copper metabolism disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having an amino acid sequence selected from the group consisting of

SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

The invention also provides a method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

## BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A, 1B and 1C show the amino acid sequence alignments between HTPH-1 (Incyte Clone number; SEQ ID NO:1) and rat ABC transporter (GI 2982567; SEQ ID NO:7), produced using the multisequence alignment program of LASERGENE™ software (DNASTAR Inc, Madison WI).

Figure 2 shows the amino acid sequence alignments between HTPH-2 (Incyte Clone number; SEQ ID NO:2) and E. coli copper homeostasis protein (GI 1736520; SEQ ID NO:8), produced using the multisequence alignment program of LASERGENE™ software (DNASTAR Inc, Madison WI).

The first column of Table 1 shows the polypeptide sequence identifiers, SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3. The second column shows the nucleotide sequence identifiers, SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6, of the consensus sequences which encode HTPH. The third column lists the Incyte Clone number in which nucleic acids encoding each HTPH were first identified. The fourth column lists the tissue library from which the clone was isolated. The fifth column lists the overlapping and/or extended nucleic acid sequences which were used to derive the consensus sequences of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6.

The first column of Table 2 lists the polypeptide sequence identifiers (SEQ ID

NO:1, SEQ ID NO:2, and SEQ ID NO:3). The second column shows the number of amino acid residues in each polypeptide. The third column lists the potential phosphorylation sites in each polypeptide, and the fourth column lists potential N-glycosylation sites. The fifth column lists any significant protein family signature or ligand/substrate binding motif present in each polypeptide. The sixth column indicates the identity of the protein. The seventh column describes the method of analysis or algorithm(s) used to identify the polypeptide.

The first column of Table 3 lists the nucleotide sequence identifiers (SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6). The second column lists the tissue expression of HTPH and fraction of total tissue which express each polynucleotide. The third column lists the disease class and fraction of total disease tissues that express each polynucleotide. The fourth column lists the vector used to subclone the cDNA library.

The first column of Table 4 list the nucleotide sequence identifiers (SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6). The second column lists the corresponding Incyte Clone number. The third column lists the description of the tissue library from which the clone was derived.

Table 5 summarizes the programs, algorithms, databases, and qualifying scores used to analyze HTPH. The first column of Table 5 shows the tool, program, or algorithm; the second column, the database; the third column, a brief description; and the forth column (where applicable), scores for determining the strength of a match between two sequences (the higher the value, the greater the homology).

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise.

Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

**DEFINITIONS**

"HTPH," as used herein, refers to the amino acid sequences, or variant thereof, of substantially purified HTPH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to HTPH, increases or prolongs the duration of the effect of HTPH. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HTPH.

An "allelic variant," as this term is used herein, is an alternative form of the gene encoding HTPH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HTPH, as described herein, include

those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as HTPH or a polypeptide with at least one functional characteristic of HTPH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HTPH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HTPH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent HTPH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of HTPH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms “amino acid” or “amino acid sequence,” as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, “fragments,” “immunogenic fragments,” or “antigenic fragments” refer to fragments of HTPH which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of HTPH. Where “amino acid sequence” is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification,” as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to HTPH, decreases the amount or the duration of the effect of the biological or immunological activity of HTPH. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HTPH.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HTPH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic

HTPH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" binds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HTPH or fragments of HTPH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts, e.g., NaCl, detergents, e.g.,sodium dodecyl sulfate (SDS), and other components, e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.

"Consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (The Perkin-Elmer Corp., Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid

sequence encoding HTPH, by Northern analysis is indicative of the presence of nucleic acids encoding HTPH in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HTPH.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity," as used herein, refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence

similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign™ program (DNASTAR, Inc., Madison WI). The MegAlign™ program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0t$ or $R_0t$ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters,

chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an arrangement of distinct polynucleotides arrayed on a substrate, e.g., paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The terms "element" or "array element" as used herein in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate," as it appears herein, refers to a change in the activity of HTPH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HTPH.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which, comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6 from related polynucleotide sequences. A fragment of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 and

the region of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment. Alternatively, a fragment when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding HTPH, or fragments thereof, or HTPH itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence

of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of HTPH polypeptides, as used herein, refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE™ software.

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to HTPH. This definition may also include, for

example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

**THE INVENTION**

The invention is based on the discovery of new human transport protein homologs (HTPH), the polynucleotides encoding HTPH, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, reproductive disorders, and copper metabolism disorders. In Table 1, columns 1 and 2 show the sequence identification numbers (SEQ ID NO:) of the amino acid and nucleic acid sequence, respectively. Column 3 shows the Clone ID of the Incyte Clone in which nucleic acids encoding each HTPH were first identified, and column 4, the cDNA library of this clone. Column 5 lists the SEQ ID numbers corresponding to the fragments in Column 6. Column 6 describes fragments, and shows the Incyte clones (and libraries) and shotgun sequences useful as fragments, e.g., in hybridization technologies, and which are part of the consensus nucleotide sequence of each HTPH.

The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the amino acid SEQ ID NO; column 2 shows the number of amino acid residues; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, the signature sequences, where applicable, that occur in the polypeptide, and column 6; analytical methods used to identify the polypeptide using sequence homologies and protein motifs.

The columns of Table 3 show the tissue expression of each nucleic acid sequence by northern analysis, diseases or disorders associated with this tissue expression, and the vector into which each cDNA was cloned.

Figures 1A, 1B, and 1C show a truncated alignment between HTPH-1 and the ABC transporter from rat (GI 2982567; SEQ ID NO:7). The alignment truncates the first 240 amino acids of the rat ABC transporter In particular, HTPH-1 and the rat ABC transporter share 90%

sequence identity, as well as sharing the ABC transporters family signature sequence and the ATP/GTP-binding site motif A (P-loop).

Figure 2 shows an alignment between HTPH-2 and the copper homeostasis protein CutC from E coli (GI 1736520; SEQ ID NO:8). In particular, HTPH-2 and the E. coli copper homeostasis protein share 36% sequence identity, as well as several phosphorylation sites.

The invention also encompasses HTPH variants. A preferred HTPH variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HTPH amino acid sequence, and which contains at least one functional or structural characteristic of HTPH.

The invention also encompasses polynucleotides which encode HTPH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6, which encodes an HTPH.

The invention also encompasses a variant of a polynucleotide sequence encoding HTPH. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HTPH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected fro the group consisting of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HTPH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HTPH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HTPH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode HTPH and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HTPH under

appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HTPH possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HTPH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HTPH and HTPH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HTPH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 $\mu$g/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 $\mu$g/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing and analysis are well known in the art. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE® (Amersham Pharmacia Biotech Ltd., Uppsala, Sweden), Taq polymerase (The Perkin-Elmer Corp., Norwalk, CT), thermostable T7 polymerase (Amersham Pharmacia Biotech Ltd., Uppsala, Sweden), or combinations of polymerases and proofreading exonucleases, such as those found in the ELONGASE™ amplification system (Life Technologies, Inc., Rockville, MD). Preferably, sequence preparation is automated with machines, e.g., the ABI CATALYST™ 800 (The Perkin-Elmer Corp., Norwalk, CT) or MICROLAB® 2200 (Hamilton Co., Reno, NV) systems, in combination with thermal cyclers. Sequencing can also be automated, such as by ABI PRISM™ 373 or 377 systems (The Perkin-Elmer Corp., Norwalk, CT) or the MEGABACE™ 1000 capillary electrophoresis system (Molecular Dynamics, Inc., Sunnyvale, CA). Sequences can be analyzed using computer programs and algorithms well known in the art. (See, e.g., Ausubel, supra, unit 7.7; and Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, Inc, New York, NY.)

The nucleic acid sequences encoding HTPH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et

al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO™ 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, (The Perkin-Elmer Corp., Norwalk, CT)), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HTPH may be cloned in recombinant DNA molecules that direct expression of HTPH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HTPH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HTPH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding HTPH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, HTPH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (The Perkin-Elmer Corp., Norwalk, CT). Additionally, the amino acid sequence of HTPH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.)

In order to express a biologically active HTPH, the nucleotide sequences encoding HTPH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HTPH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HTPH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HTPH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where

only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HTPH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding HTPH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HTPH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HTPH can be achieved using a multifunctional E. coli vector such as Bluescript® (Stratagene) or pSport1™ plasmid (GIBCO BRL). Ligation of sequences encoding HTPH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HTPH are needed, e.g. for the production of antibodies, vectors which direct high level expression of HTPH may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HTPH. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-54; Scorer, C. A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of HTPH. Transcription of sequences encoding HTPH may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HTPH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HTPH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

For long term production of recombinant proteins in mammalian systems, stable expression of HTPH in cell lines is preferred. For example, sequences encoding HTPH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable

marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* or *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP) (Clontech, Palo Alto, CA), ß glucuronidase and its substrate ß-D-glucuronoside, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HTPH is inserted within a marker gene sequence, transformed cells containing sequences encoding HTPH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HTPH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HTPH and that express HTPH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HTPH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques

include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HTPH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; Coligan, J. E. et al. (1997 and periodic supplements) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HTPH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HTPH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HTPH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HTPH may be designed to contain signal sequences which direct secretion of HTPH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from

the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HTPH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HTPH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HTPH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HTPH encoding sequence and the heterologous protein sequence, so that HTPH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10. A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled HTPH may be achieved in vitro using the TNT™ rabbit reticulocyte lysate or wheat germ extract systems (Promega, Madison, WI). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, preferably $^{35}$S-methionine.

Fragments of HTPH may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (The Perkin-Elmer Corp., Norwalk, CT). Various fragments of HTPH may be synthesized separately and then combined to produce the full length molecule.

**THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between HTPH-1 and the ABC transporter from rat (GI 2982567). In addition, protein sequence analysis (PFAM and BLOCKS) identify HTPH-1 as an ABC transporter (ABC_tran, and BL00211, respectively). HTPH-1 is expressed in cancerous libraries, and in libraries from reproductive tissues. Therefore, HTPH-1 appears to play a role in cancer and reproductive disorders. Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between HTPH-2 and the copper homeostasis protein CutC from E. coli (GI 1736520). In addition, HTPH-2 is expressed in cancerous libraries, and in libraries from reproductive tissues. Therefore, HTPH-2 appears to play a role in cancer, reproductive disorders, and copper metabolism disorders. Protein sequence analysis identifies HTPH-3 as a LacY family proton/sugar symporter ((BL00896). In addition, HTPH-3 is expressed in cancerous libraries, and in libraries from reproductive tissues. Therefore, HTPH-3 appears to play a role in cancer and reproductive disorders.

Therefore, in one embodiment, HTPH or a fragment or derivative thereof may be administered to a subject to treat or prevent a copper metabolism disorder. Such disorders can include, but are not limited to, Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX.

In another embodiment, a vector capable of expressing HTPH or a fragment or derivative thereof may be administered to a subject to treat or prevent a copper metabolism disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HTPH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a copper metabolism disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HTPH may be administered to a subject to treat or prevent a copper metabolism disorder including, but not limited to, those listed above.

In a further embodiment, an antagonist of HTPH may be administered to a subject to treat or prevent a cancer. Such a cancer may include, but is not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds HTPH may be used directly as an antagonist or indirectly as a

targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HTPH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HTPH may be administered to a subject to treat or prevent a cancer including, but not limited to, those described above.

In a further embodiment, an antagonist of HTPH may be administered to a subject to treat or prevent a reproductive. Such a disorder may include, but is not limited to, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, carcinoma of the male breast, and gynecomastia. In one aspect, an antibody which specifically binds HTPH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HTPH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding HTPH may be administered to a subject to treat or prevent a disorder including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects. This is of particular importance with respect to agents that regulate the activity of HTPH-1. HTPH-1 is a member of the ABC transporter family, which includes such members as the P-glycoprotein, MDR (the multidrug-resistance protein). These proteins mediate the transport of drugs, e.g., antibiotics, out of target cells, thereby reducing the efficacy of the drug in the target cell. Agents that prevent the removal of drug from the target cell will therefore significantly enhance the efficacy of that drug, and as such, are of particular importance in combination therapies.

An antagonist of HTPH may be produced using methods which are generally known in the

art. In particular, purified HTPH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HTPH. Antibodies to HTPH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of polyclonal antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HTPH or with any fragment or oligopeptide thereof which has immunogenic properties. Rats and mice are preferred hosts for downstream applications involving monoclonal antibody production. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable. (For review of methods for antibody production and analysis, see, e.g., Harlow, E. and Lane, D. (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.)

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HTPH have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 14 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of HTPH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to HTPH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and

Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HTPH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HTPH may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity and minimal cross-reactivity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between HTPH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HTPH epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HTPH. Affinity is expressed as an association constant, $K_a$, which is defined as the molar concentration of HTPH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The $K_a$ determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HTPH epitopes, represents the average affinity, or avidity, of the antibodies for HTPH. The $K_a$ determined for a preparation of monoclonal antibodies, which are monospecific for a particular HTPH epitope, represents a true measure of affinity. High-affinity antibody preparations with $K_a$ ranging from about $10^9$ to $10^{12}$ L/mole are preferred for use in immunoassays in which the HTPH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with $K_a$ ranging from about $10^6$ to $10^7$ L/mole are preferred for use in immunopurification and similar procedures which ultimately require

dissociation of HTPH, preferably in active form, from the antibody. (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, D. C.; and Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York, NY.)

The titre and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of HTPH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding HTPH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding HTPH may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HTPH. Thus, complementary molecules or fragments may be used to modulate HTPH activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HTPH.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HTPH. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding HTPH can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HTPH. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HTPH. Oligonucleotides derived from the transcription

initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HTPH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HTPH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms

of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HTPH, antibodies to HTPH, and mimetics, agonists, antagonists, or inhibitors of HTPH. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active

compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HTPH, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HTPH or fragments thereof, antibodies of HTPH, and agonists, antagonists or inhibitors of HTPH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the $ED_{50}$ (the dose therapeutically effective in 50% of the population) or $LD_{50}$ (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the $ED_{50}/LD_{50}$ ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the $ED_{50}$ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account

include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind HTPH may be used for the diagnosis of disorders characterized by expression of HTPH, or in assays to monitor patients being treated with HTPH or agonists, antagonists, or inhibitors of HTPH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HTPH include methods which utilize the antibody and a label to detect HTPH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HTPH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HTPH expression. Normal or standard values for HTPH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HTPH under conditions suitable for complex formation The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HTPH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HTPH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be

used to detect and quantitate gene expression in biopsied tissues in which expression of HTPH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HTPH, and to monitor regulation of HTPH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HTPH or closely related molecules may be used to identify nucleic acid sequences which encode HTPH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding HTPH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably have at least 50% sequence identity to any of the HTPH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6 or from genomic sequences including promoters, enhancers, and introns of the HTPH gene.

Means for producing specific hybridization probes for DNAs encoding HTPH include the cloning of polynucleotide sequences encoding HTPH or HTPH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as $^{32}P$ or $^{35}S$, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HTPH may be used for the diagnosis of a disorder associated with expression of HTPH. Examples of such a disorder include, but are not limited to, copper metabolism disorders such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX, cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, and reproductive disorders such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic

pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding HTPH may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered HTPH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HTPH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HTPH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HTPH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HTPH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HTPH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may

provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HTPH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HTPH, or a fragment of a polynucleotide complementary to the polynucleotide encoding HTPH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HTPH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HTPH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial

P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HTPH on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, HTPH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HTPH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with HTPH, or fragments thereof, and washed. Bound HTPH is then detected by methods well known in the art. Purified HTPH can also be coated directly onto plates

for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HTPH specifically compete with a test compound for binding HTPH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HTPH.

In additional embodiments, the nucleotide sequences which encode HTPH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 09/113,427, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from CLONTECH Laboratories, Inc. (Palo Alto, CA) or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL™ (Life Technologies, Inc., Rockville, MD), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega Corp., Madison, WI), OLIGOTEX™ latex particles (QIAGEN Inc., Valencia, CA), or an OLIGOTEX™ mRNA purification kit (QIAGEN Inc., Valencia, CA). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE™ mRNA purification kit (Ambion, Austin, TX).

In some cases, Stratagene, Inc. (La Jolla, CA), was provided with RNA and constructed

the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP™ vector system (Stratagene, Inc., La Jolla, CA) or SUPERSCRIPT™ plasmid system (Life Technologies, Inc., Rockville, MD), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, supra, 1997, units 5.1-6.6) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL® S1000, SEPHAROSE® CL2B, or SEPHAROSE® CL4B column chromatography (Amersham Pharmacia Biotech Ltd., Uppsala, Sweden) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., pBLUESCRIPT™ (Stratagene, Inc., La Jolla, CA), pSPORT™ 1 (Life Technologies, Inc., Rockville, MD), or pINCY (Incyte Pharmaceuticals, Inc., Palo Alto, CA). Recombinant plasmids were transformed into competent E. coli cells, e.g., the XL1-Blue, XL1-BlueMRF, or SOLR™ strains (Stratagene, Inc., La Jolla, CA), or DH5α™, DH10B, or ElectroMAX DH10B (Life Technologies, Inc., Rockville, MD).

**II. Isolation of cDNA Clones**

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP™ vector system (Stratagene, Inc., La Jolla, CA), or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD® Minipreps DNA purification system (Promega Corp., Madison, WI); an AGTC® Miniprep purification kit (Edge Biosystems, Gaithersburg, MD); the QIAWELL® 8 Plasmid, QIAWELL® 8 Plus Plasmid, or the QIAWELL® 8 Ultra Plasmid purification systems (QIAGEN Inc., Valencia, CA); or the R.E.A.L.™ Prep 96 plasmid kit (QIAGEN Inc., Valencia, CA). Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format. (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN® dye (Molecular Probes, Inc., Eugene, OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

**III. Sequencing and Analysis**

The cDNAs were prepared for sequencing using either an ABI PRISM™ CATALYST™ 800 (Perkin-Elmer Applied Biosystems, Foster City, CA) or a MICROLAB® 2200 (Hamilton Co., Reno, NV) sequencing preparation system in combination with Peltier PTC-200 thermal

cyclers (MJ Research, Inc., Watertown, MA). The cDNAs were sequenced using the ABI PRISM™ 373 or 377 sequencing systems and ABI protocols, base calling software, and kits (Perkin-Elmer Applied Biosystems). Alternatively, solutions and dyes from Amersham Pharmacia Biotech, Ltd. were used in place of the ABI kits. In some cases, reading frames were determined using standard methods (Ausubel, supra). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the software programs used, corresponding algorithms, references, and cutoff parameters used where applicable. The references cited in the third column of Table 5 are incorporated by reference herein. Sequence alignments were also analyzed and produced using MACDNASIS PRO software (Hitachi Software Engineering Co., Ltd. San Bruno, CA) and the multisequence alignment program of LASERGENE software (DNASTAR Inc, Madison WI).

The polynucleotide sequences were validated by removing vector, linker, and polyA tail sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. This was followed by translation of the full length polynucleotide sequences to derive the corresponding full length amino acid sequences. These full length polynucleotide and amino acid sequences were subsequently analyzed by querying against databases such as the GenBank databases described above and SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

**IV. Northern Analysis**

Northern analysis is a laboratory technique used to detect the presence of a transcript and involves the hybridization of a labeled nucleotide probe to a substrate on which mRNAs from a particular cell type or tissue have been bound (Sambrook, supra, ch. 7).

Electronic northern analysis was performed using BLAST to search for identical or related sequences in nucleotide databases such as the LIFESEQ® database (Incyte Pharmaceuticals) or GenBank. The sensitivity of the computer search was modified to set the specificity of the match. The basis of the search was the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score encompassed both the degree of similarity and the length of match between the two sequences. For example, with a product score of 40, the match is exact within a 1% to 2% error, and, with a product score of 70, the match is exact. Related molecules show product scores between 15 and 40, although lower scores may identify related sequences if the match is biased toward 5' sequence or nonconserved portions of the sequences being examined.

Electronic northern analysis involved the grouping of cDNA libraries into organ/tissue and disease categories. The organ/tissue categories are cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic, and the disease categories are cancer, inflammation/trauma, fetal, neurological, and pooled. The libraries in each category which expressed the sequence were counted, and the fraction of the total was calculated. The results of electronic Northern analysis were reported as the percentage distribution of the sequence in the organ/tissue and disease categories in Table3.

**V. Extension of HTPH Encoding Polynucleotides**

Full-length nucleic acid sequences (SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6) were produced by extension of the component fragments described in Table 1, Column 5, using oligonucleotide primers based on those fragments. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO™ 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR™ kit (The Perkin-Elmer Corp., Norwalk, CT) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the PTC-200 thermal cycler (MJ Research, Inc., Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1 94° C for 1 min (initial denaturation)

| | |
|---|---|
| Step 2 | 65° C for 1 min |
| Step 3 | 68° C for 6 min |
| Step 4 | 94° C for 15 sec |
| Step 5 | 65° C for 1 min |
| Step 6 | 68° C for 7 min |
| Step 7 | Repeat steps 4 through 6 for an additional 15 cycles |
| Step 8 | 94° C for 15 sec |
| Step 9 | 65° C for 1 min |
| Step 10 | 68° C for 7:15 min |
| Step 11 | Repeat steps 8 through 10 for an additional 12 cycles |
| Step 12 | 72° C for 8 min |
| Step 13 | 4° C (and holding) |

A 5 $\mu$l to 10 $\mu$l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK™ (QIAGEN Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 $\mu$l of ligation buffer, 1$\mu$l T4-DNA ligase (15 units) and 1$\mu$l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in 40 $\mu$l of appropriate media) were transformed with 3 $\mu$l of ligation mixture and cultured in 80 $\mu$l of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37°C, the E. coli mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150 $\mu$l of liquid LB/2x carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 $\mu$l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 $\mu$l from each sample was transferred into a PCR array.

For PCR amplification, 18 $\mu$l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

| | |
|---|---|
| Step 1 | 94° C for 60 sec |
| Step 2 | 94° C for 20 sec |
| Step 3 | 55° C for 30 sec |
| Step 4 | 72° C for 90 sec |
| Step 5 | Repeat steps 2 through 4 for an additional 29 cycles |
| Step 6 | 72° C for 180 sec |
| Step 7 | 4° C (and holding) |

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequences of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

**VI. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO™ 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 $\mu$Ci of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex™ G-25 superfine size exclusion dextran bead column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing $10^7$ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

**VII. Microarrays**

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the

scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

**VIII. Complementary Polynucleotides**

Sequences complementary to the HTPH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HTPH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO™ 4.06 software and the coding sequence of HTPH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HTPH-encoding transcript.

**IX. Expression of HTPH**

Expression and purification of HTPH is achieved using bacterial or virus-based expression systems. For expression of HTPH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express HTPH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HTPH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HTPH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription.

Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HTPH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Pharmacia, Piscataway, NJ). Following purification, the GST moiety can be proteolytically cleaved from HTPH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak, Rochester, NY). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN Inc, Chatsworth, CA). Methods for protein expression and purification are discussed in Ausubel, F. M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch 10, 16. Purified HTPH obtained by these methods can be used directly in the following activity assay.

**X. Demonstration of HTPH Activity**

HTPH transport activity can be demonstrated through the use of a ligand mixing assay that is used to measure transport from early to late endosomal compartments in X. laevis oocytes. Ovaries are dissected from adult female X. laevis, and oocytes are isolated. (Mukhopadhyay A, et al. (1997) J. Cell. Biol. 136(6): 1227-1237). Oocytes are pulsed with 2mg/ml avidin for 5hrs at 18° C, washed, then incubated for 16 hrs to allow avidin to transport to a late compartment. The oocytes are then incubated with 1mg/ml biotin-horseradish peroxidase (HRP) for 30 minutes at 18° C to label early endocytic compartments. Varying amounts of HTPH are injected into the oocytes, and the oocytes are incubated at 18° C. Oocytes are collected at several time points after HTPH injection, washed, and lysed in 100µl of phosphate-buffered saline containing 0.3% Triton X-100, 0.2% methylbenzethorium chloride, and 400 µg/ml of BSA-biotin as a scavenger. Finally, the lysates are centrifuged for 30 seconds in a microfuge, and the avidin-biotin complexes are immunoprecipitated using anti-avidin antibody-coated plates by incubation at 4 °C overnight. The plates are washed at least 5 times to remove unbound proteins. Transport from the early endosomes to the late compartments is quantified by measuring the amount of immunoprecipitated HRP; increased transport due to HTPH is quantitated by comparison with control oocytes.

## XI. Functional Assays

HTPH function is assessed by expressing the sequences encoding HTPH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT™ (Life Technologies, Gaithersburg, MD) and pCR™ 3.1 (Invitrogen, Carlsbad, CA, both of which contain the cytomegalovirus promoter. 5-10 $\mu$g of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 $\mu$g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate properties, for example, their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York, NY.

The influence of HTPH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HTPH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success, NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HTPH and other genes of interest can be analyzed by Northern analysis or microarray techniques.

## XII. Production of HTPH Specific Antibodies

HTPH substantially purified using polyacrylamide gel electrophoresis (PAGE)(see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is

used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HTPH amino acid sequence is analyzed using LASERGENE™ software (DNASTAR Inc.) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XIII. Purification of Naturally Occurring HTPH Using Specific Antibodies**

Naturally occurring or recombinant HTPH is substantially purified by immunoaffinity chromatography using antibodies specific for HTPH. An immunoaffinity column is constructed by covalently coupling anti-HTPH antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HTPH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HTPH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HTPH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HTPH is collected.

**XIV. Identification of Molecules Which Interact with HTPH**

HTPH, or biologically active fragments thereof, are labeled with $^{125}$I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HTPH, washed, and any wells with labeled HTPH complex are assayed. Data obtained using different concentrations of HTPH are used to calculate values for the number, affinity, and association of HTPH with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred

embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

**TABLE 1**

| Protein SEQ ID NO: | Nucleotide SEQ ID NO: | Clone ID | Library | Fragments |
|---|---|---|---|---|
| 1 | 4 | 2074412 | ISLTNOT01 | 624415R1 (PGANNOT01), 624415X12 (PGANNOT01), 624415X15 (PGANNOT01), 2048538T6 (LIVRFET02), 2074412H1 (ISLTNOT01). |
| 2 | 5 | 2704671 | PONSAZT01 | 531314F1 (BRAINOT03), 1341229F1 (COLNTUT03), 1753719F6 (LIVRTUT01), 2704671H1 (PONSAZT01), 2885141H1 (SINJNOT02), 3128067H1 (LUNGTUT12). |
| 3 | 6 | 3094754 | CERVNOT03 | 078177R1 (SYNORAB01), 137825F1 (SYNORAB01), 137825R1 (SYNORAB01), 2463208H1 (THYRNOT08), 3094754H1 (CERVNOT03). |

# TABLE 2

| Protein SEQ ID NO: | Amino Acid Residues | Potential Phosphorylation Sites | Potential glycosylation sites | Signature Sequence | Identification | Analytical Methods |
|---|---|---|---|---|---|---|
| 1 | 574 | T126 S330 T337 T411 S460 S487 T491 T563 T26 T73 S74 S89 T194 S393 | N179 N230 N409 N507 | ABC: L459-I473<br>P loop: G355-S362 | ABC transporter | BLAST, PFAM, BLOCKS |
| 2 | 261 | S80 S160 S165 S215 S235 S8 S17 S80 T193 S212 S215 S240 | N225 | | Copper transporter | BLAST |
| 3 | 163 | S56 S149 | N77 N142 | L27-R73 | Sugar transporter | BLOCKS |

## TABLE 3

| Nucleotide SEQ ID NO: | Tissue Expression (Fraction of Total) | Disease Class (Fraction of Total) | Vector |
|---|---|---|---|
| 4 | Reproductive (0.364) Nervous (0.182) Developmental (0.121) | Cancer (0.455) Fetal (0.182) Inflammation (0.182) | pINCY |
| 5 | Reproductive (0.314) Gastrointestinal (0.137) Nervous (0.118) | Cancer (0.549) Inflammation (0.157) Trauma (0.157) | pINCY |
| 6 | Reproductive (0.256) Cardiovascular (0.137) Gastrointestinal (0.128) | Cancer (0.496) Inflammation (0.222) Fetal (0.179) | pINCY |

# TABLE 4

| Nucleotide SEQ ID NO: | Clone ID | Library | Library Comment |
|---|---|---|---|
| 4 | 2074412 | ISLTNOT01 | Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells. |
| 5 | 2704671 | PONSAZT01 | Library was constructed using RNA isolated from diseased pons tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease. |
| 6 | 3094754 | CERVNOT03 | Library was constructed using RNA isolated from uterine cervical tissue removed from a 40-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy and dilation and curettage. Pathology indicated secretory phase endometrium. |

**TABLE 5**

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ABI FACTURA | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences. | Perkin-Elmer Applied Biosystems, Foster City, CA. | |
| ABI/PARACEL FDF | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences. | Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA. | Mismatch <50% |
| ABI AutoAssembler | A program that assembles nucleic acid sequences. | Perkin-Elmer Applied Biosystems, Foster City, CA. | |
| BLAST | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx. | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402. | *ESTs:* Probability value= 1.0E-8 or less<br>*Full Length sequences:* Probability value= 1.0E-10 or less |
| FASTA | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489. | *ESTs:* fasta E value=1.06E-6<br>*Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less<br>*Full Length sequences:* fastx score=100 or greater |
| BLIMPS | A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions. | Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424. | Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less |
| PFAM | A Hidden Markov Models-based application useful for protein family search. | Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322. | Score=10-50 bits, depending on individual protein families |

**TABLE 5 (cont.)**

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. | Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221. | Score= 4.0 or greater |
| Phred | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability. | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194. | |
| Phrap | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA. | Score= 120 or greater; Match length= 56 or greater |
| Consed | A graphical tool for viewing and editing Phrap assemblies | Gordon, D. et al. (1998) Genome Res. 8:195-202. | |
| SPScan | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439. | Score=5 or greater |
| Motifs | A program that searches amino acid sequences for patterns that matched those defined in Prosite. | Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI. | |

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof.

2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide sequence of claim 3.

7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, SEQ ID N O:6, and fragments thereof..

8. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 7.

9. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 7.

10. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

11. A host cell comprising the expression vector of claim 10.

12. A method for producing a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof, the method comprising the steps of:

a) culturing the host cell of claim 11 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

13. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

14. A purified antibody which specifically binds to the polypeptide of claim 1.

15. A purified agonist of the polypeptide of claim 1.

16. A purified antagonist of the polypeptide of claim 1.

17. A method for treating or preventing a cancer, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

18. A method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

19. A method for treating or preventing a copper metabolism disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 13.

20. A method for detecting a polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, and fragments thereof in a biological sample, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the

hybridization complex correlates with the presence of the polynucleotide encoding the polypeptide in the biological sample.

21. The method of claim 20 further comprising amplifying the polynucleotide prior to hybridization.

```
1   - - - - - - - - - - - - - - - - - - - - - - - - - - - - M G  2074412
241 D L G R K L R L L S G Y L W P R G S P S L Q L T V L L C M G  GI 2982567

3   L M G V R R A L N V F V P I F Y R N I V N L L T E K A P W N  2074412
271 L M G L D R A L N V L V P I F Y R D I V N L L T S K A P W S  GI 2982567

33  S L A W T V T S Y V F L K F L Q G G G T G S T G F V S N L R  2074412
301 S L A W T V T T Y V F L K F L Q G G G T G S T G F V S N L R  GI 2982567

63  T F L W I R V Q Q F T S R R V E L L I F S H L H E L S L R W  2074412
331 T F L W I R V Q Q F T S R G V E L R L F S H L H E L S L R W  GI 2982567

93  H L G R R T G E V L R I A D R G T S S V T G L L S Y L V F N  2074412
361 H L G R R T G E V L R I V D R G T S S V T G L L S Y L V F N  GI 2982567

123 V I P T L A D I I I G I I Y F S M F F N A W F G L I V F L C  2074412
391 I I P T L A D I I I G I I Y F S M F F N A W F G L I V F L C  GI 2982567

153 M S L Y L T L T I V V T E W R T K F R R A M N T Q E N A T R  2074412
421 M S L Y L I L T I M V T E W R A K F R R D M N T Q E N A T R  GI 2982567

183 A R A V D S L L N F E T V K Y Y N A E S Y E V E R Y R E A I  2074412
451 A R A V D S L L N F E T V K Y Y N A E G Y E L E R Y R E A I  GI 2982567

213 I K Y Q G L E W K S S A S L V L L N Q T Q N L V I G L G L L  2074412
481 L K F Q G L E W K S T A S L V L L N Q T Q N M V I G F G L L  GI 2982567
```

FIGURE 1A

```
243 A G S L L C A Y F V T E Q K L Q V G D Y V L F G T Y I I Q L 2074412
511 A G S L L C A Y F V S E R R L Q V G D F V L F G T Y I T Q L GI 2982567

273 Y M P L N W F G T Y Y R M I Q T N F I D M E N M F D L L K E 2074412
541 Y M P L N W F G T Y Y R M I Q T N F I D M E N M F D L L K E GI 2982567

303 E T E V K D L P G A G P L R F Q K G R I E F E N V H F S Y A 2074412
571 E T E V K D V P G A G P L R F H K G R V E F E N V H F S Y A GI 2982567

333 D G R E T L Q D V S F T V M P G Q T L A L V G P S G A G K S 2074412
601 D G R E T L Q D V S F T V M P G Q T V A L V G P S G A G K S GI 2982567

363 T I L R L L F R F Y D I S S G C I R I D G Q D I S Q V T Q A 2074412
631 T I L R L L F R F Y D I S S G C I R I D G Q D I S Q V T Q I GI 2982567

393 S L R S H I G V V P Q D T V L F N D T I A D N I R Y G R V T 2074412
661 S L R S H I G V V P Q D T V L F N D T I A N N I R Y G R V T GI 2982567

423 A G N D E V E A A A Q A A G I H D A I M A F P E G Y R T Q V 2074412
691 A G D S E I Q A A A Q A A G I H D A I L S F P E G Y E T Q V GI 2982567

453 G E R G L K L S G G E K Q R V A I A R T I L K A P G I I L L 2074412
721 G E R G L K L S G G E K Q R V A I A R T I L K A P D I I L L GI 2982567

483 D E A T S A L D T S N E R A I Q A S L A K V C A N R T T I V 2074412
751 D E A T S A L D T S N E R A I Q A S L A K V C T N R T T I V GI 2982567
```

FIGURE 1B

```
513 V A H R L S T V V N A D Q I L V I K D G C I V E R G R H E A  2074412
781 V A H R L S T V V N A D Q I L V I K D G C I I E R G R H E A  GI 2982567

543 L L S R G G V Y A D M W Q L Q Q G Q E E T S E D T K P Q T M  2074412
811 L L S R G G V Y A E M W Q L Q Q Q G Q E T V P E - - - - - -  GI 2982567

573 E R                                                          2074412
835 D S                                                          GI 2982567
```

FIGURE 1C

```
1    MKRQGASSERKRARIPSGKAGAANGFLMEV 2704671
1    MALLEICCYSMECALTAQQNGADR---VEL GI 1736520

31   CVDSVESAVNAERGEGGTTPSMGVLQVVKQ 2704671
28   C-----------AAPKEGGLTPSLGVLKSVRQ GI 1736520

61   SVQIPVFVMIRPRGGDFLYSDREIEVMKAD 2704671
49   RVTIPVHPIIRPRGGDFCYSDGEFAAILED GI 1736520

91   IRLAKLYGADGLVFGALTEDGHIDKELCMS 2704671
79   VRTVRELGFPGLVTGVLDVDGNVDMPRMEK GI 1736520

121  LMAICRPLPVTFHRAFDMVHDPMAALETLL 2704671
109  IMAAAGPLAVTFHRAFDMCANPLYTLNNLA GI 1736520

151  TLGFERVLTSGCDSSALEGLPLIKRLIEQA 2704671
139  ELGIARVLTSGQKSDALQGLSKIMELIAH- GI 1736520

181  KGRIVVMPGGGITDRNLQRILEGSGATEFH 2704671
168  RDAPIIMAGAGVRAENLHHFLDA-GVLEVH GI 1736520

211  CSARSTRDSGMKFRNSSVAMGASLSCSEYS 2704671
197  SSAGAWQASPMRYRNQGLSMSSDEHADEYS GI 1736520

241  LKVTDVTKVRTLNAI-AKNILV 2704671
227  RYIVDGAAVAEMKGIIERHQAK GI 1736520
```

FIGURE 2

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
HILLMAN, Jennifer L.
YUE, Henry
REDDY, Roopa
GORGONE, Gina
CORLEY, Neil C.
AZIMZAI, Yalda
PATTERSON, Chandra
BAUGHN, Mariah R.

<120> HUMAN TRANSPORT PROTEIN HOMOLOGS

<130> PF-0555 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/113,427
<151> 1998-07-10

<160> 8

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 574
<212> PRT
<213> HOMO SAPIENS

<220>
<221> misc_feature
<223> Incyte Clone No: 2074412, ISLTNOT01

<300>

<400> 1
Met Gly Leu Met Gly Val Arg Arg Ala Leu Asn Val Phe Val Pro Ile
1 5 10 15
Phe Tyr Arg Asn Ile Val Asn Leu Leu Thr Glu Lys Ala Pro Trp Asn
20 25 30
Ser Leu Ala Trp Thr Val Thr Ser Tyr Val Phe Leu Lys Phe Leu Gln
35 40 45
Gly Gly Gly Thr Gly Ser Thr Gly Phe Val Ser Asn Leu Arg Thr Phe
50 55 60
Leu Trp Ile Arg Val Gln Gln Phe Thr Ser Arg Arg Val Glu Leu Leu
65 70 75 80
Ile Phe Ser His Leu His Glu Leu Ser Leu Arg Trp His Leu Gly Arg
85 90 95
Arg Thr Gly Glu Val Leu Arg Ile Ala Asp Arg Gly Thr Ser Ser Val
100 105 110
Thr Gly Leu Leu Ser Tyr Leu Val Phe Asn Val Ile Pro Thr Leu Ala
115 120 125
Asp Ile Ile Ile Gly Ile Ile Tyr Phe Ser Met Phe Phe Asn Ala Trp
130 135 140
Phe Gly Leu Ile Val Phe Leu Cys Met Ser Leu Tyr Leu Thr Leu Thr
145 150 155 160
Ile Val Val Thr Glu Trp Arg Thr Lys Phe Arg Arg Ala Met Asn Thr
165 170 175
Gln Glu Asn Ala Thr Arg Ala Arg Ala Val Asp Ser Leu Leu Asn Phe

180 185 190
Glu Thr Val Lys Tyr Tyr Asn Ala Glu Ser Tyr Glu Val Glu Arg Tyr
195 200 205
Arg Glu Ala Ile Ile Lys Tyr Gln Gly Leu Glu Trp Lys Ser Ser Ala
210 215 220
Ser Leu Val Leu Leu Asn Gln Thr Gln Asn Leu Val Ile Gly Leu Gly
225 230 235 240
Leu Leu Ala Gly Ser Leu Leu Cys Ala Tyr Phe Val Thr Glu Gln Lys
245 250 255
Leu Gln Val Gly Asp Tyr Val Leu Phe Gly Thr Tyr Ile Ile Gln Leu
260 265 270
Tyr Met Pro Leu Asn Trp Phe Gly Thr Tyr Tyr Arg Met Ile Gln Thr
275 280 285
Asn Phe Ile Asp Met Glu Asn Met Phe Asp Leu Leu Lys Glu Glu Thr
290 295 300
Glu Val Lys Asp Leu Pro Gly Ala Gly Pro Leu Arg Phe Gln Lys Gly
305 310 315 320
Arg Ile Glu Phe Glu Asn Val His Phe Ser Tyr Ala Asp Gly Arg Glu
325 330 335
Thr Leu Gln Asp Val Ser Phe Thr Val Met Pro Gly Gln Thr Leu Ala
340 345 350
Leu Val Gly Pro Ser Gly Ala Gly Lys Ser Thr Ile Leu Arg Leu Leu
355 360 365
Phe Arg Phe Tyr Asp Ile Ser Ser Gly Cys Ile Arg Ile Asp Gly Gln
370 375 380
Asp Ile Ser Gln Val Thr Gln Ala Ser Leu Arg Ser His Ile Gly Val
385 390 395 400
Val Pro Gln Asp Thr Val Leu Phe Asn Asp Thr Ile Ala Asp Asn Ile
405 410 415
Arg Tyr Gly Arg Val Thr Ala Gly Asn Asp Glu Val Glu Ala Ala Ala
420 425 430
Gln Ala Ala Gly Ile His Asp Ala Ile Met Ala Phe Pro Glu Gly Tyr
435 440 445
Arg Thr Gln Val Gly Glu Arg Gly Leu Lys Leu Ser Gly Gly Glu Lys
450 455 460
Gln Arg Val Ala Ile Ala Arg Thr Ile Leu Lys Ala Pro Gly Ile Ile
465 470 475 480
Leu Leu Asp Glu Ala Thr Ser Ala Leu Asp Thr Ser Asn Glu Arg Ala
485 490 495
Ile Gln Ala Ser Leu Ala Lys Val Cys Ala Asn Arg Thr Thr Ile Val
500 505 510
Val Ala His Arg Leu Ser Thr Val Val Asn Ala Asp Gln Ile Leu Val
515 520 525
Ile Lys Asp Gly Cys Ile Val Glu Arg Gly Arg His Glu Ala Leu Leu
530 535 540
Ser Arg Gly Gly Val Tyr Ala Asp Met Trp Gln Leu Gln Gln Gly Gln
545 550 555 560
Glu Glu Thr Ser Glu Asp Thr Lys Pro Gln Thr Met Glu Arg
565 570

<210> 2
<211> 261
<212> PRT
<213> HOMO SAPIENS

<220>
<221> misc_feature
<223> Incyte Clone No: 2704671, PONSAZT01

<300>

<400> 2
Met Lys Arg Gln Gly Ala Ser Ser Glu Arg Lys Arg Ala Arg Ile Pro
1 5 10 15

Ser Gly Lys Ala Gly Ala Ala Asn Gly Phe Leu Met Glu Val Cys Val
20 25 30
Asp Ser Val Glu Ser Ala Val Asn Ala Glu Arg Gly Glu Gly Gly Thr
35 40 45
Thr Pro Ser Met Gly Val Leu Gln Val Val Lys Gln Ser Val Gln Ile
50 55 60
Pro Val Phe Val Met Ile Arg Pro Arg Gly Gly Asp Phe Leu Tyr Ser
65 70 75 80
Asp Arg Glu Ile Glu Val Met Lys Ala Asp Ile Arg Leu Ala Lys Leu
85 90 95
Tyr Gly Ala Asp Gly Leu Val Phe Gly Ala Leu Thr Glu Asp Gly His
100 105 110
Ile Asp Lys Glu Leu Cys Met Ser Leu Met Ala Ile Cys Arg Pro Leu
115 120 125
Pro Val Thr Phe His Arg Ala Phe Asp Met Val His Asp Pro Met Ala
130 135 140
Ala Leu Glu Thr Leu Leu Thr Leu Gly Phe Glu Arg Val Leu Thr Ser
145 150 155 160
Gly Cys Asp Ser Ser Ala Leu Glu Gly Leu Pro Leu Ile Lys Arg Leu
165 170 175
Ile Glu Gln Ala Lys Gly Arg Ile Val Val Met Pro Gly Gly Gly Ile
180 185 190
Thr Asp Arg Asn Leu Gln Arg Ile Leu Glu Gly Ser Gly Ala Thr Glu
195 200 205
Phe His Cys Ser Ala Arg Ser Thr Arg Asp Ser Gly Met Lys Phe Arg
210 215 220
Asn Ser Ser Val Ala Met Gly Ala Ser Leu Ser Cys Ser Glu Tyr Ser
225 230 235 240
Leu Lys Val Thr Asp Val Thr Lys Val Arg Thr Leu Asn Ala Ile Ala
245 250 255
Lys Asn Ile Leu Val
260

<210> 3
<211> 163
<212> PRT
<213> HOMO SAPIENS

<220>
<221> misc_feature
<223> Incyte Clone No: 3094754, CERVNOT03

<300>

<400> 3
Met Gly Ser Thr Trp Gly Ser Pro Gly Trp Val Arg Leu Ala Leu Cys
1 5 10 15
Leu Thr Gly Leu Val Leu Ser Leu Tyr Ala Leu His Val Lys Ala Ala
20 25 30
Arg Ala Arg Asp Arg Asp Tyr Arg Ala Leu Cys Asp Val Gly Thr Ala
35 40 45
Ile Ser Cys Ser Arg Val Phe Ser Ser Arg Trp Gly Arg Gly Phe Gly
50 55 60
Leu Val Glu His Val Leu Gly Gln Asp Ser Ile Leu Asn Gln Ser Asn
65 70 75 80
Ser Ile Phe Gly Cys Ile Phe Tyr Thr Leu Gln Leu Leu Leu Gly Cys
85 90 95
Leu Arg Thr Arg Trp Ala Ser Val Leu Met Leu Leu Ser Ser Leu Val
100 105 110
Ser Leu Ala Gly Ser Val Tyr Leu Ala Trp Ile Leu Phe Phe Val Leu
115 120 125
Tyr Asp Phe Cys Ile Val Cys Ile Thr Thr Tyr Ala Ile Asn Val Ser
130 135 140

Leu Met Trp Leu Ser Phe Arg Lys Val Gln Glu Pro Gln Gly Lys Ala
 145 150 155 160
 Lys Arg His

<210> 4
<211> 1933
<212> DNA
<213> HOMO SAPIEN

<220>
<221> misc_feature
<223> Incyte Clone No: 2074412, ISLTNOT01

<300>

<400> 4
gggtattatc accttcgtct gacctagatg gggctcatgg gtgtccgacg ggcactcaat 60
gtgtttgtgc ctatattcta taggaacatt gtgaacttgc tgactgagaa ggcaccttgg 120
aactctctgg cctggactgt taccagttac gtcttcctca agttcctcca ggggggtggc 180
actggcagta caggcttcgt gagcaacctg cgcaccttcc tgtggatccg ggtgcagcag 240
ttcacgtctc ggcgggtgga gctgctcatc ttctcccacc tgcacgagct ctcactgcgc 300
tggcacctgg ggcgccgcac aggggaggtg ctgcggatcg cggatcgggg cacatccagt 360
gtcacagggc tgctcagcta cctggtgttc aatgtcatcc ccacgctggc cgacatcatc 420
attggcatca tctacttcag catgttcttc aacgcctggt ttggcctcat tgtgttcctg 480
tgcatgagtc tttacctcac cctgaccatt gtggtcactg agtggagaac caagtttcgt 540
cgtgctatga acacacagga gaacgctacc cgggcacgag cagtggactc tctgctaaac 600
ttcgagacgg tgaagtatta caacgccgag agttacgaag tggaacgcta tcgagaggcc 660
atcatcaaat atcagggttt ggagtggaag tcgagcgctt cactggtttt actaaatcag 720
acccagaacc tggtgattgg gctcgggctc ctcgccggct ccctgctttg cgcatacttt 780
gtcactgagc agaagctaca ggttggggac tatgtgctct ttggcaccta cattatccag 840
ctgtacatgc ccctcaattg gtttggcacc tactacagga tgatccagac caacttcatt 900
gacatggaga acatgtttga cttgctgaaa gaggagacag aagtgaagga ccttcctgga 960
gcagggcccc ttcgctttca gaagggccgt attgagtttg agaacgtgca cttcagctat 1020
gccgatgggc gggagactct gcaggacgtg tctttcactg tgatgcctgg acagacactt 1080
gccctggtgg gcccatctgg ggcagggaag agcacaattt tgcgcctgct gtttcgcttc 1140
tacgacatca gctctggctg catccgaata gatgggcagg acatttcaca ggtgacccag 1200
gcctctctcc ggtctcacat tggagttgtg ccccaagaca ctgtcctctt taatgacacc 1260
atcgccgaca atatccgtta cggccgtgtc acagctggga atgatgaggt ggaggctgct 1320
gctcaggctg caggcatcca tgatgccatt atggctttcc ctgaagggta caggacacag 1380
gtgggcgagc ggggactgaa gctgagcggc ggggagaagc agcgcgtcgc cattgcccgc 1440
accatcctca aggctccggg catcattctg ctggatgagg caacgtcagc gctggataca 1500
tctaatgaga gggccatcca ggcttctctg gccaaagtct gtgccaaccg caccaccatc 1560
gtagtggcac acaggctctc aactgtggtc aatgctgacc agatcctcgt catcaaggat 1620
ggctgcatcg tggagagggg acgacacgag gctctgttgt cccgaggtgg ggtgtatgct 1680
gacatgtggc agctgcagca gggacaggaa gaaacctctg aagacactaa gcctcagacc 1740
atggaacggt gacaaaagtt tggccacttc cctctcaaag actaacccag aagggaataa 1800
gatgtgtctc ctttccctgg cttatttcat cctggtcttg ggtatggtg ctagctatgg 1860
taagggaaag ggacctttcc gaaaaacatc ttttggggaa ataaaaatgt ggactgtgcg 1920
aggaaaaaaa aaa 1933

<210> 5
<211> 1358
<212> DNA
<213> HOMO SAPIENS

<220>
<221> misc_feature
<223> Incyte Clone No: 2704671, PONSAZT01

<300>

<400> 5
cagaagtgcg cctgcgttgc gactccgctt tcgcgggggc tgctggggcg taggtgtcgg 60

```
ggacgcgcgc acgggcgcgc gcagctgttg acgcgcttct tagctggtgc gcgccggagc      120
ccaaattcca agtggaaact gcaggcgcac gagggaggaa cgcgtggagc atgaaaaggc      180
aggggcctc ctctgagcga aaacgagcgc ggataccgtc cgggaaggcc ggagcagcaa      240
atggatttct catggaagtt tgtgttgatt cagtggaatc agctgtgaat gcagaaagag      300
gagaggggg aactacaccc agcatgggtg tccttcaagt agtgaagcag agtgttcaga      360
tcccagtttt tgtgatgatt cggccacggg gaggtgattt tttgtattca gatcgtgaaa      420
ttgaggtgat gaaggctgac attcgtcttg ccaagcttta tggtgctgat ggtttggttt      480
ttggggcatt gactgaagat ggacacattg acaaagagct gtgtatgtcc cttatggcta      540
tttgccgccc tctgccagtc actttccacc gagcctttga catggttcat gatccaatgg      600
cagctctgga gaccctctta accttgggat ttgaacgcgt gttgaccagt ggatgtgaca      660
gttcagcatt agaagggcta cccctaataa agcgactcat tgagcaggca aaaggcagga      720
ttgtggtaat gccaggaggt ggtataacag acagaaatct acaaaggatc cttgagggtt      780
caggtgctac agaattccac tgttctgctc ggtctactag agactcggga atgaagtttc      840
gaaattcatc tgttgccatg ggagcctcac tttcttgctc agaatattcc ctaaaggtaa      900
cagatgtgac caaagtaagg actttgaatg ctatcgcaaa gaacatcctg gtgtagccag      960
acctctctga gagacatgga tatcacagga tgaaggtaga actataatct gcaattctct     1020
atgacacagc tttaaccttc ttctctggcc aggacagtcg caatctttgt tttaagtttc     1080
acatggccat ggagaatgtg cccaagaaga aaaagaattt gaaacagaga tacagtcact     1140
tcctttgctt agtcttacca gtgattgtca tcatggttaa agctggtctg tgcttcttcc     1200
atagacagaa gcttagtctg ttttcagtgg aattaattga tgaactggga aaattttaac     1260
tgcatggtat gaattcagag tgtgacttaa gggtcaattc aaagcagtat tttgactttt     1320
catttgtaaa ataaaaattt ccactattaa aaaaaaa                              1358
```

<210> 6
<211> 1045
<212> DNA
<213> HOMO SAPIEN

<220>
<221> misc_feature
<223> Incyte Clone No: 3094754, CERVNOT03

<300>

<400> 6

```
ctcgagccgt gcctggcagc cgctggggcc gcgattccgc acgtccctta cccgcttcac       60
tagtcccggc attcttcgct gttttcctaa ctcgcccgct tgactagcgc cctggaacag      120
ccatttgggt cgtggagtgc gagcacggcc ggccaatcgc cgagtcagag ggccaggagg      180
ggcgcggcca ttcgccgccc ggcccctgct ccgtggctgg ttttctccgc gggcgcctcg      240
ggcggaacct ggagataatg ggcagcacct gggggagccc tggctgggtg cggctcgctc      300
tttgcctgac gggcttagtg ctctcgctct acgcgctgca cgtgaaggcg gcgcgcgccc      360
gggaccggga ttaccgcgcg ctctgcgacg tgggcaccgc catcagctgt tcgcgcgtct      420
tctcctccag gtggggcagg ggtttcgggc tggtggagca tgtgctggga caggacagca      480
tcctcaatca atccaacagc atattcggtt gcatcttcta cacactacag ctattgttag      540
gttgcctgcg gacacgctgg gcctctgtcc tgatgctgct gagctccctg gtgtctctcg      600
ctggttctgt ctacctggcc tggatcctgt tcttcgtgct ctatgatttc tgcattgttt      660
gtatcaccac ctatgctatc aacgtgagcc tgatgtggct cagtttccgg aaggtccaag      720
aaccccaggg caaggctaag aggcactgag ccctcaaccc aagccaggct gacctcatct      780
gctttgcttt ggcatgtgag ccttgcctaa gggggcatat ctgggtccct agaaggccct      840
agatgtgggg cttctagatt acccccctcct cctgccatac ccgcacatga caatggacca      900
aatgtgccac acgctcgctc ttttttacac ccagtgcctc tgactctgtc cccatgggct      960
ggtctccaaa gctctttcca ttgcccaggg agggaaggtt ctgagcaata aagtttctta     1020
gatcaatcaa aaaaaaaaaa aaaaa                                           1045
```

<210> 7
<211> 836
<212> PRT
<213> RATTUS NORVEGICUS

<300>
<308> g2982567

<400> 7

Met Val Thr Val Gly Asn Tyr Cys Glu Ala Glu Gly Pro Ala Gly Pro

```
1               5                   10                  15
Ala Trp Thr Gln Asn Gly Leu Ser Pro Cys Phe Phe Tyr Thr Leu Val
            20                  25                  30
Pro Ser Thr Leu Met Thr Leu Gly Val Leu Ala Leu Val Leu Val Leu
        35                  40                  45
Pro Cys Arg Arg Arg Glu Val Pro Ala Gly Thr Glu Glu Leu Ser Trp
    50                  55                  60
Ala Ala Gly Pro Arg Val Ala Pro Tyr Ala Leu Gln Leu Ser Leu Ala
65                  70                  75                  80
Ile Leu Gln Met Ala Leu Pro Leu Ala Ser Leu Ala Gly Arg Val Gly
                85                  90                  95
Thr Ala Arg Gly Val Arg Leu Pro Gly Tyr Leu Leu Leu Ala Ser Val
            100                 105                 110
Leu Glu Ser Leu Ala Ser Ala Cys Gly Leu Trp Leu Leu Val Val Glu
        115                 120                 125
Arg Ser Gln Ala Arg Gln Ser Leu Ala Met Gly Val Trp Met Lys Phe
    130                 135                 140
Arg His Ser Leu Gly Leu Leu Leu Leu Trp Thr Val Thr Phe Ala Ala
145                 150                 155                 160
Glu Asn Leu Val Leu Val Ser Trp Asn Ser Pro Gln Trp Trp Trp Ser
                165                 170                 175
Arg Ala Asp Leu Gly Gln Gln Val Gln Phe Gly Leu Trp Val Leu Arg
            180                 185                 190
Tyr Met Thr Ser Gly Gly Leu Phe Ile Leu Gly Leu Trp Ala Pro Gly
        195                 200                 205
Leu Arg Pro Gln Ser Tyr Thr Leu His Val Asn Glu Glu Asp Gln Asp
    210                 215                 220
Gly Gly Arg Asn Gln Gly Arg Ser Thr Asp Pro Arg Ser Thr Trp Arg
225                 230                 235                 240
Asp Leu Gly Arg Lys Leu Arg Leu Leu Ser Gly Tyr Leu Trp Pro Arg
                245                 250                 255
Gly Ser Pro Ser Leu Gln Leu Thr Val Leu Leu Cys Met Gly Leu Met
            260                 265                 270
Gly Leu Asp Arg Ala Leu Asn Val Leu Val Pro Ile Phe Tyr Arg Asp
        275                 280                 285
Ile Val Asn Leu Leu Thr Ser Lys Ala Pro Trp Ser Ser Leu Ala Trp
    290                 295                 300
Thr Val Thr Thr Tyr Val Phe Leu Lys Phe Leu Gln Gly Gly Gly Thr
305                 310                 315                 320
Gly Ser Thr Gly Phe Val Ser Asn Leu Arg Thr Phe Leu Trp Ile Arg
                325                 330                 335
Val Gln Gln Phe Thr Ser Arg Gly Val Glu Leu Arg Leu Phe Ser His
            340                 345                 350
Leu His Glu Leu Ser Leu Arg Trp His Leu Gly Arg Arg Thr Gly Glu
        355                 360                 365
Val Leu Arg Ile Val Asp Arg Gly Thr Ser Ser Val Thr Gly Leu Leu
    370                 375                 380
Ser Tyr Leu Val Phe Asn Ile Ile Pro Thr Leu Ala Asp Ile Ile Ile
385                 390                 395                 400
Gly Ile Ile Tyr Phe Ser Met Phe Phe Asn Ala Trp Phe Gly Leu Ile
                405                 410                 415
Val Phe Leu Cys Met Ser Leu Tyr Leu Ile Leu Thr Ile Met Val Thr
            420                 425                 430
Glu Trp Arg Ala Lys Phe Arg Arg Asp Met Asn Thr Gln Glu Asn Ala
        435                 440                 445
Thr Arg Ala Arg Ala Val Asp Ser Leu Leu Asn Phe Glu Thr Val Lys
    450                 455                 460
Tyr Tyr Asn Ala Glu Gly Tyr Glu Leu Glu Arg Tyr Arg Glu Ala Ile
465                 470                 475                 480
Leu Lys Phe Gln Gly Leu Glu Trp Lys Ser Thr Ala Ser Leu Val Leu
                485                 490                 495
Leu Asn Gln Thr Gln Asn Met Val Ile Gly Phe Gly Leu Leu Ala Gly
            500                 505                 510
```

Ser Leu Leu Cys Ala Tyr Phe Val Ser Glu Arg Arg Leu Gln Val Gly
515 520 525
Asp Phe Val Leu Phe Gly Thr Tyr Ile Thr Gln Leu Tyr Met Pro Leu
530 535 540
Asn Trp Phe Gly Thr Tyr Tyr Arg Met Ile Gln Thr Asn Phe Ile Asp
545 550 555 560
Met Glu Asn Met Phe Asp Leu Leu Lys Glu Glu Thr Glu Val Lys Asp
565 570 575
Val Pro Gly Ala Gly Pro Leu Arg Phe His Lys Gly Arg Val Glu Phe
580 585 590
Glu Asn Val His Phe Ser Tyr Ala Asp Gly Arg Glu Thr Leu Gln Asp
595 600 605
Val Ser Phe Thr Val Met Pro Gly Gln Thr Val Ala Leu Val Gly Pro
610 615 620
Ser Gly Ala Gly Lys Ser Thr Ile Leu Arg Leu Leu Phe Arg Phe Tyr
625 630 635 640
Asp Ile Ser Ser Gly Cys Ile Arg Ile Asp Gly Gln Asp Ile Ser Gln
645 650 655
Val Thr Gln Ile Ser Leu Arg Ser His Ile Gly Val Val Pro Gln Asp
660 665 670
Thr Val Leu Phe Asn Asp Thr Ile Ala Asn Asn Ile Arg Tyr Gly Arg
675 680 685
Val Thr Ala Gly Asp Ser Glu Ile Gln Ala Ala Ala Gln Ala Ala Gly
690 695 700
Ile His Asp Ala Ile Leu Ser Phe Pro Glu Gly Tyr Glu Thr Gln Val
705 710 715 720
Gly Glu Arg Gly Leu Lys Leu Ser Gly Gly Glu Lys Gln Arg Val Ala
725 730 735
Ile Ala Arg Thr Ile Leu Lys Ala Pro Asp Ile Ile Leu Leu Asp Glu
740 745 750
Ala Thr Ser Ala Leu Asp Thr Ser Asn Glu Arg Ala Ile Gln Ala Ser
755 760 765
Leu Ala Lys Val Cys Thr Asn Arg Thr Thr Ile Val Val Ala His Arg
770 775 780
Leu Ser Thr Val Val Asn Ala Asp Gln Ile Leu Val Ile Lys Asp Gly
785 790 795 800
Cys Ile Ile Glu Arg Gly Arg His Glu Ala Leu Leu Ser Arg Gly Gly
805 810 815
Val Tyr Ala Glu Met Trp Gln Leu Gln Gln Gln Gly Gln Glu Thr Val
820 825 830
Pro Glu Asp Ser
835

<210> 8
<211> 248
<212> PRT
<213> ESCHERICHIA COLI

<300>
<308> g1736520

<400> 8
Met Ala Leu Leu Glu Ile Cys Cys Tyr Ser Met Glu Cys Ala Leu Thr
1 5 10 15
Ala Gln Gln Asn Gly Ala Asp Arg Val Glu Leu Cys Ala Ala Pro Lys
20 25 30
Glu Gly Gly Leu Thr Pro Ser Leu Gly Val Leu Lys Ser Val Arg Gln
35 40 45
Arg Val Thr Ile Pro Val His Pro Ile Ile Arg Pro Arg Gly Gly Asp
50 55 60
Phe Cys Tyr Ser Asp Gly Glu Phe Ala Ala Ile Leu Glu Asp Val Arg
65 70 75 80
Thr Val Arg Glu Leu Gly Phe Pro Gly Leu Val Thr Gly Val Leu Asp
85 90 95

```
Val Asp Gly Asn Val Asp Met Pro Arg Met Glu Lys Ile Met Ala Ala
            100                 105                 110
Ala Gly Pro Leu Ala Val Thr Phe His Arg Ala Phe Asp Met Cys Ala
        115                 120                 125
Asn Pro Leu Tyr Thr Leu Asn Asn Leu Ala Glu Leu Gly Ile Ala Arg
    130                 135                 140
Val Leu Thr Ser Gly Gln Lys Ser Asp Ala Leu Gln Gly Leu Ser Lys
145                 150                 155                     160
Ile Met Glu Leu Ile Ala His Arg Asp Ala Pro Ile Ile Met Ala Gly
                165                 170                     175
Ala Gly Val Arg Ala Glu Asn Leu His His Phe Leu Asp Ala Gly Val
            180                 185                 190
Leu Glu Val His Ser Ser Ala Gly Ala Trp Gln Ala Ser Pro Met Arg
        195                 200                 205
Tyr Arg Asn Gln Gly Leu Ser Met Ser Ser Asp Glu His Ala Asp Glu
    210                 215                 220
Tyr Ser Arg Tyr Ile Val Asp Gly Ala Ala Val Ala Glu Met Lys Gly
225                 230                 235                     240
Ile Ile Glu Arg His Gln Ala Lys
                245
```